

2021-03

Assessment of hepatic steatosis by controlled attenuation parameter using the M and XL probes: an individual patient data meta-analysis

Petroff, D

<http://hdl.handle.net/10026.1/18283>

10.1016/s2468-1253(20)30357-5

The Lancet Gastroenterology & Hepatology

Elsevier

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

Controlled attenuation parameter including the XL probe to assess steatosis: an individual patient data meta-analysis

David Petroff PhD^{1,2*}, Valentin Blank MD^{2,3*}, Prof Philip N. Newsome MD^{4,5,6}, Shalimar MD⁷, Cosmin
5 Sebastian Voican MD^{8,9,10}, Prof Maja Thiele MD¹¹, Prof Victor de Lédinghen MD¹², Stephan Baumeler MD¹³,
Prof Wah Kheong Chan MD¹⁴, Prof Gabriel Perlemuter MD^{8,9,10}, Ana-Carolina Cardoso MD¹⁵, Prof Sandeep
Aggarwal MD¹⁶, Magali Sasso PhD¹⁷, Peter J. Eddowes MD^{4,5,6,18}, Michael Allison MD¹⁹, Emmanouil
Tsochatzis MD²⁰, Prof Quentin M. Anstee MD²¹, David Sheridan MD²², Jeremy F. Cobbold MD²³, Prof Sylvie
10 Naveau MD^{8,9,10}, Monica Lupsor-Platon MD²⁴, Prof Sebastian Mueller MD²⁵, Prof Aleksander Krag MD¹¹,
Marie Irlès-Depe MD¹², David Semela MD¹³, Prof Grace Lai-Hung Wong MD^{26,27}, Prof Vincent Wai-Sun Wong
MD^{26,27}, Prof Cristiane A. Villela-Nogueira MD¹⁵, Harshit Garg MD²⁸, Prof Olivier Chazouillères MD²⁹,
Johannes Wiegand MD³⁰, Thomas Karlas MD³

1 University of Leipzig, Clinical Trial Centre, Leipzig, Germany
2 University of Leipzig, Faculty of Medicine, Integrated Research and Treatment Center (IFB) AdiposityDiseases, Leipzig,
15 Germany
3 Division of Gastroenterology, Department of Medicine II, Leipzig University Medical Center, Leipzig, Germany
4 National Institute for Health Research (NIHR), Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS
Foundation Trust, University of Birmingham, Birmingham, United Kingdom
5 Centre for Liver Research, Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, United
20 Kingdom
6 Liver Unit, University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom
7 Department of Gastroenterology, All India Institute of Medical Sciences, New Delhi, India
8 Faculté de Médecine Paris-Sud, Univ Paris-Sud, Université Paris-Saclay, Le Kremlin-Bicêtre, France
9 Service d'Hépatogastroentérologie et Nutrition, Hôpital Antoine-Béclère, Hôpitaux Universitaires Paris-Sud, Assistance
25 Publique-Hôpitaux de Paris, Clamart, France
10 INSERM U996, DHU Hepatinov, Labex LERMIT, Clamart, France
11 Department of Gastroenterology and Hepatology, Odense University Hospital, University of Southern Denmark, Odense,
Denmark
12 Centre d'Investigation de la Fibrose Hépatique, Bordeaux University Hospital, Pessac, France; INSERM U1053, Bordeaux
30 University, Bordeaux, France
13 Department of Gastroenterology/Hepatology, Cantonal Hospital St Gallen, Switzerland
14 Gastroenterology and Hepatology Unit, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur,
Malaysia
15 Hepatology Unit, Hospital Universitário Clementino Fraga Filho, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.
35 16 Department of Surgical Disciplines, All India Institute of Medical Sciences, New Delhi, India
17 R & D Department, Echosens, Paris, France
18 National Institute for Health Research Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and
University of Nottingham, Nottingham, UK
19 Liver Unit, Addenbrooke's Hospital, Cambridge Biomedical Research Centre, Cambridge, UK
40 20 University College London Institute for Liver and Digestive Health, Royal Free Hospital, London, UK
21 Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK; &
Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne,
UK
22 Institute of Translational and Stratified Medicine, Faculty of Medicine and Dentistry, University of Plymouth, Plymouth, UK
45 23 Department of Gastroenterology and Hepatology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital,
Oxford, UK.
24 Department of Medical Imaging, Iuliu Hatieganu University of Medicine and Pharmacy, Regional Institute of Gastroenterology
and Hepatology, Cluj-Napoca, Romania.
25 Department of Medicine and Liver Diseases, Salem Medical Center, University of Heidelberg, Germany.
50 26 Institute of Digestive Disease, The Chinese University of Hong Kong, Sha Tin, Hong Kong
27 Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Sha Tin, Hong Kong
28 Department of Surgical Disciplines, All India Institute of Medical Sciences, New Delhi, India
29 Hepatology department, Saint-Antoine Hospital, Assistance Publique – Hôpitaux de Paris (APHP), Centre de Recherche Saint-
Antoine (CRSA), Sorbonne University, Paris, France
55 30 Division of Hepatology, Department of Medicine II, Leipzig University Medical Center, Leipzig, Germany

* DP and VB contributed equally to this work.

60 **Short title:** IPDMA on CAP for steatosis including the XL probe

Corresponding author

Thomas Karlas, M.D.
Leipzig University Medical Center
Division of Gastroenterology
65 Department of Medicine II
Liebigstraße 20
04103 Leipzig
Germany

70 Tel: +49 341 97 12240
Fax: +49 341 97 12209
Email: thomas.karlas@medizin.uni-leipzig.de

Keywords

75 Controlled attenuation parameter (CAP); vibration controlled transient elastography (VCTE); liver steatosis; non-alcoholic fatty liver disease (NAFLD)

Abbreviations

AUC: Area under the curve

80 ALD: Alcoholic liver disease

AST: Aspartate transaminase

BMI: Body mass index

CAP: Controlled attenuation parameter

CI: confidence interval

85 IPDMA: Individual patient data meta-analysis

IQR: Interquartile range

LSM: Liver stiffness measurement

MRI: Magnetic resonance imaging

NAFLD: Non-alcoholic fatty liver disease

90 NAS: NAFLD Activity Score

NASH: Non-alcoholic steatohepatitis

NPV: Negative predictive value

PPV: Positive predictive value

ROC: Receiver operating characteristics

95 SLD: Skin-to-liver-capsule distance

VCTE: Vibration controlled Transient elastography

3 Figures and 3 Tables

100 **Registration**

Prospero CRD42018099284

RESEARCH IN CONTEXT

Evidence before this study

105 Original research and meta-analyses showed that non-invasive grading of hepatic steatosis with the ultrasound based controlled attenuation parameter (CAP) is accurate over a variety of aetiologies when using the so-called M probe. A limitation of this probe is that it is not appropriate for patients with obesity meaning that the important application to non-alcoholic fatty liver disease (NAFLD) was limited. Recently, the XL probe was introduced, which was designed for patients with higher body mass index (BMI). A number of studies were
110 conducted using CAP with the XL probe and showed promising results for steatosis grading compared to the histological gold-standard. Consensus regarding optimal cut-offs was not available and uncertainties in diagnostic performance and relevant covariates led us to conduct an individual patient data meta-analysis. In a systematic literature search, we identified 16 studies with histology-controlled CAP including the XL probe. We were able to obtain individual data from 13 of them and 2346 patients were included in the final analysis.

115 Added value of this study

The diagnostic performance and optimal cut-offs for CAP with the XL probe depend substantially on aetiology, diabetes status, BMI and sex. The area under the curve (AUC) from receiver operating characteristic (ROC) analysis is only 0.754 and 0.717 for detection of higher grade steatosis (S2-3 or S3, respectively) in NAFLD patients. The AUC for detection of any steatosis (S1-3) is better at 0.819, but S0 cases are heavily influenced by
120 patients scheduled for bariatric surgery. Overall, the diagnostic performance of CAP in bariatric surgery patients is better than for the average NAFLD patient. Previously proposed reliability criteria do not lead to improved performance. Comparison of M and XL probe measurements in overweight patients where both probes are viable shows high variance for the difference between measurements though the mean difference is small.

Implications of all the available evidence

125 The interpretation of CAP must take aetiology and covariates into account. Even with the XL probe, CAP cannot grade steatosis in patients with NAFLD adequately. The findings in this study should be taken into account when developing or interpreting non-invasive diagnosis of NASH with methods that include CAP. Its value in a NAFLD screening setting needs to be studied, ideally with methods beyond the traditional histological reference standard.

130 SUMMARY

Background

Diagnostic tools for liver disease now include estimation of steatosis (S0-S3) and controlled attenuation parameter (CAP) is one such non-invasive candidate. It has become available for obese patients (FibroScan XL probe), but consensus is lacking regarding cut-offs and its diagnostic performance. This individual patient data meta-analysis aims to assess diagnostic properties and identify relevant covariates.

Methods

Sixteen studies reported histology-controlled CAP including the XL probe, and individual data from 13 papers and 2346 patients were included. Probe recommendation was based on automated selection, manual assessment of skin-to-liver-capsule distance and a body mass index (BMI) criterion. Receiving operating characteristic methods and mixed models were used to assess diagnostic properties and covariates. Patients with non-alcoholic fatty liver disease (NAFLD) were analysed separately because of their predominant requirement for use of the XL probe (registration: CRD42018099284).

Findings

Patients (age 46 ± 14 years) were recruited from 20 centres in nine countries (29%, 23% and 47% were normal weight, overweight and obese respectively; 54%, 18%, 12% and 13% had NAFLD, viral hepatitis, alcohol associated liver disease and other aetiologies). The XL probe was recommended in 1050 patients, 930 of whom (89%) had NAFLD, and the areas under the curve were 0.82 (S0 vs S1-3; 95% CI 0.77 to 0.87) and 0.75 (S0-1 vs S2-3; 95% CI 0.72 to 0.79). CAP was independently affected by aetiology, diabetes, BMI, aspartate transaminase and sex. Optimal cut-offs differed substantially across aetiology.

Interpretation

CAP cut-offs vary according to aetiology, and effectively recognize significant steatosis in patients with viral hepatitis. CAP cannot grade steatosis in patients with NAFLD adequately, but its value in a NAFLD screening setting needs to be studied, ideally with methods beyond the traditional histological reference standard.

155 Funding

Supported in part by the German Federal Ministry of Education and Research and Echosens, which provided funding, but did not influence study design or analysis.

INTRODUCTION

Fatty liver, the hepatic manifestation of the metabolic syndrome, has grown to become a major field of clinical research, and the development of fast and easily applicable non-invasive methods for assessment of fatty liver disease is of great importance.¹ Both serum-based markers and liver stiffness measurement (LSM) have evolved to become established and valuable tools to assess the degree of hepatic fibrosis.^{1,2} More recently, non-invasive quantification of hepatic steatosis has attracted scientific interest and the continued development of technologies will help define the clinical implications of steatosis.³⁻⁷ Magnetic resonance imaging (MRI) methods permit precise estimation of total hepatic fat, which is of value in clinical studies,⁸ but is not ideal for the screening of large populations at risk.⁹ On the other hand, the ultrasound-based controlled attenuation parameter (CAP) technology allows estimation of hepatic fat content during LSM with vibration-controlled transient elastography (VCTE) on the FibroScan equipment (Echosens, Paris, France).¹⁰ CAP was originally available only with the M probe (appropriate for lean subjects and skin-to-liver-capsule distance (SLD) of less than 25 mm) and discriminates fairly well between histological grades of steatosis in patients with viral hepatitis and NAFLD, as shown in an individual patient data meta-analysis (IPDMA) performed by our team.¹¹ The CAP algorithm later introduced for the XL probe facilitates steatosis assessment in obese patients (SLD \geq 25 mm),¹² where steatosis assessment is most needed. Several biopsy controlled studies have assessed correlation of XL-CAP with histological steatosis grades, especially in obese patients with non-alcoholic fatty liver disease (NAFLD).¹²⁻²⁷

The availability of both probes may provide the practitioner with the necessary tools to grade steatosis quickly and non-invasively in patients at risk, as recently demonstrated in one multi-centred prospective study on patients undergoing liver biopsy for suspicion of NAFLD or non-alcoholic steatohepatitis (NASH).¹⁴ However, an in-depth meta-analysis considering all relevant covariates and reliability criteria is lacking, which is necessary to account for cohorts with mixed aetiology and broader inclusion criteria. Moreover, larger cohorts are needed to evaluate diagnostic performance and ascertain whether or not cut-offs established with the M probe are appropriate for the XL probe, especially considering that the M probe may have been used inappropriately in overweight patients.²⁸ The introduction of the XL probe is intended to open up the CAP field to obese patients, where steatosis quantification is most relevant. To assess its performance and provide a comprehensive comparison with the M probe, our research team performed an extensive IPDMA from studies that included use of the XL probe.

METHODS

This study was approved by the ethics committee of the coordinating site in Leipzig (218/15-ek) and the study protocol was registered before start of the data analysis (Prospero CRD42018099284).

Search strategy and study selection criteria

190 Search terms and time intervals were defined at two consensus meetings of the study group in Paris 2018 and Vienna 2019. Only published or accepted English language manuscripts by a peer-reviewed journal until April 30th 2019 qualified for inclusion in the meta-analysis. Study results published only in abstract form were not considered.

Publications on the diagnostic properties of CAP for liver steatosis quantification with histology as reference
195 standard were extracted from electronic databases from their inception until April 30th 2019 including “PubMed” and “Web of Science”. The systematic literature search was performed using the term “controlled attenuation parameter XL”. The final query for PubMed was: "controlled"[All Fields] AND "attenuation"[All Fields] AND "parameter"[All Fields] AND "XL"[All Fields] AND "english"[Language] NOT
200 "review"[Publication Type] AND 1900/1/1:2019/4/30[Date - Entry]. An analogous query was used for Web of Science. Titles and abstract were screened for eligibility by two authors (TK and VB). In case of disagreement, a third researcher (DP) was consulted to arrive at a consensus. The articles selected for a full text review were examined by two researchers (TK, DP) to determine whether or not the inclusion/exclusion criteria were met.

In addition to this search strategy, we screened abstract books of the International Liver Congress (EASL, 2016-2019) and The Liver Meeting (AASLD, 2016-2018) using the same search terms as above. The corresponding
205 authors were contacted and invited to participate in the meta-analysis if their full papers were accepted within the pre-defined time limit.

Inclusion and exclusion criteria

Studies reporting original biopsy-controlled data of CAP for non-invasive grading of steatosis were eligible for inclusion in the IPDMA. CAP for the VCTE XL probe, which became commercially available in 2014, had to be
210 included in the paper. Histology specimens must have been evaluated for steatosis according to the percentage of hepatocytes involved.²⁹ Approval of the local ethics committee was required and had to be forwarded on to the analysis team. Studies that selected patients based primarily on the result of a CAP measurement were not included in the IPDMA to avoid selection bias. Studies were also excluded if data could not be obtained despite

multiple attempts to contact the study investigators. If individual data sets were published repeatedly, only the latest data set was used.

Individual data verification and study quality assessment

The quality of each study was appraised by two authors (VB and TK for half the articles and VB and DP for the other half) using the “Quality Assessment of Diagnostic Accuracy Studies 2” (QUADAS 2).³⁰ Discrepancies were discussed and resolved by these three authors. As an important further test of data quality and to ensure that they were correctly interpreted, individual data from each study were used to reproduce the main results. Inconsistencies were communicated to the corresponding authors of the studies and resolved via discussions and revised data sets, as necessary.

Steatosis grading and fibrosis staging

Histopathological data were used as reference standard for the evaluation of CAP. Steatosis was defined according to the number of affected hepatocytes: S0 (<5%), S1 (5–33%), S2 (34–66%), S3 (>66%).²⁹ Fibrosis was staged according to the Metavir (for viral hepatitis) or NAFLD Activity Score (NAS) (for NAFLD and alcoholic liver disease (ALD)).^{29, 31} Quality criteria for the histological specimens (i.e. number of portal tracts and length of specimen) differed between studies, however, the risk of sampling variability is low for steatosis grading.³² We therefore excluded only samples that were classified as unreliable by the respective study pathologist. Because the steatosis grade can be modulated by lifestyle and medical interventions, individual patients were excluded if the time interval between biopsy and CAP measurement was >6 months, as stipulated in the protocol and ratified at the study group meetings mentioned above.

Vibration-controlled transient elastography including CAP measurement

CAP acquisition took place in all studies during LSM with the VCTE device Fibroscan in line with recent guideline recommendations.^{2, 33} Cases with at least 10 single measurements for calculation of the CAP median were required. The manufacturer recommends using SLD at the measuring site for choice of the appropriate probe and has incorporated an automated probe selection software in the latest model of the device. However, different methods of probe selection have been proposed^{33, 2} and were expected amongst the studies. Therefore, we used an approach to define “correct” probe choice based on availability of data. These were automated probe selection, manual assessment of SLD, but also a body mass index (BMI) criterion proposed in international guidelines (cut-off: 30 kg/m²).² If anthropometric data indicated a high risk of inappropriate probe application,

the respective cases were excluded from the primary analysis: For the XL probe, extreme cases with BMI < 18 kg/m² were not considered, and M probe was not accepted in cases with BMI > 35 kg/m² unless the probe was selected by the internal software algorithm or SLD was <25 mm.³⁴ In all analyses of diagnostic performance, the data from the correct probe were always used. Following recommendations on the interpretation of LSM data for fibrosis estimation,³⁵ reliability criteria for CAP based on the interquartile range (IQR) of the single CAP measurements have recently been proposed: Wong et al.³⁶ suggested an IQR-threshold >40 dB/m for poorly reliable measurements with the M probe, whereas Semmler et al.³⁷ reported the relation of CAP-IQR/median of <0.1, <0.2, and <0.3 as a reliability indicator for both M and XL probe. We therefore analyzed the ability of these criteria to identify reliable CAP measurements.

Objectives

The primary objective was to establish optimal cut-offs distinguishing mild (\leq S1) from advanced steatosis (S2-S3) for the XL probe and healthy (S0) from affected (S1-3) patients, if the number of S0 patients was sufficiently high. Secondary objectives were to (i) estimate the probability of a given steatosis grade at any specific CAP value, (ii) estimate the effects of covariates on CAP values, (iii) examine the use of established cut-offs, (iv) compare M and XL probes and (v) evaluate CAP quality criteria.

Data analysis

All analyses were performed using the software R, version 3.6.1. Receiver operating characteristic (ROC) analyses including area under the curve (AUC) made use of the package “OptimalCutpoints” and the cut-off found using the median Youden optimum from a bootstrap method with 10,000 samples.³⁸ For comparison with the literature, the rule-in/rule-out optimal cut-offs were chosen by requiring that their respective sensitivity (rule-in)/specificity (rule-out) be 0.9. Confidence intervals for variables based on proportions were constructed using Wilson score intervals³⁹ and for the optimal cut-points, sensitivity and specificity from the quantiles of the bootstrapping samples. As stipulated in the protocol, a linear mixed model for CAP values included steatosis grade, BMI, diabetes status, sex, choice of probe and aetiology as covariates and the study as a random term and confidence intervals were determined with a profiling method.⁴⁰ These models were compared using ANOVA to the analogous fixed model (without study or with study as a fixed term). This permitted assessment of heterogeneity between studies from the variance of the random term, the coefficients in the study terms of the fixed effects model and the difference in the remaining coefficients between models with and without study terms. Ordinal regression from the “MASS” package⁴⁰ is used to estimate the probability of each steatosis grade

at a given CAP value after sampling the largest possible number of patients from the data with a prescribed prevalence of steatosis grades. Comparison of the M and XL probes was carried out with Bland-Altman methods. Two suggestions for CAP quality criteria were evaluated, one suggesting that the interquartile range (IQR) of 10 CAP measurements be smaller than 40 dB/m³⁶ and the other that the ratio of IQR to the median value be below 0.1, 0.2, or 0.3.³⁷ These criteria were examined using the linear correlation coefficient between CAP and histologically determined percentage of affected hepatocytes at different quality thresholds.

Role of the funding source

DP, TK and JW received an unrestricted research grant from Echosens, Paris, France, for this research. Echosens had no role in study design, collection, analysis of data, preparation of the manuscript draft. Prior to submission of the manuscript, Echosens was presented a copy that could be commented on, but there was no obligation to abide by any suggestions. The co-authors had final say on all aspects of the manuscript. This article was not commissioned by Echosens or any other company. DP, VB and TK had full access to all of the data and the final responsibility to submit for publication.

285 **RESULTS**

Papers and Patients

Figure 1 shows that the original search yielded 78 papers of which 16 met the study criteria; see **Supplementary table 1** for details about the studies. We were able to correspond with authors from all papers, but three did not provide data because of concerns on the part of the funding US agency regarding data sharing.

290 The remaining 13 papers comprised 2664 patients, 2346 of whom met the inclusion criteria for the current analysis (88%). Note that due to overlapping patients, two papers are listed in the following only as Chan 2018.^{26,22} Two papers provided more than 300 patients, six between 100 and 200 and the remaining four under 100 patients. CAP was performed within one day of the reference test in 1717 (73%) and within one month in 2292 (97%) patients.

295 Patient characteristics can be found in **Table 1** and an assessment of the quality of the studies according to QUADAS-2³⁰ can be found in **Supplementary figure 1** and shows that the risk of bias was relatively low, but that patient selection and conduct of the index test could have introduced bias when applied to our research question.

Of note is the high proportion of patients with NAFLD (54%) and subjects with high BMI, many of whom were 300 candidates for bariatric surgery. However, in the NAFLD cohort, only 14% of the patients had a BMI greater than or equal to 45 kg/m². Interestingly, 96/1277 (13%) patients with NAFLD were classified as S0, 72 of which were being considered for bariatric surgery and 17 came from a screening setting. There was a high prevalence of significant fibrosis ($F > 1$, 42%) which was unevenly distributed between those patients with (38%) and without NAFLD patients (48%). Among the former, those with S0 did not have advanced fibrosis. Note that the 305 LSM values in Table 1 based on the M and XL probes cannot be compared, since they derive largely from different patients.

Table 1: Population characteristics. ALD = Alcoholic liver disease; ALT = alanine transaminase; AST = aspartate aminotransferase; BMI = body mass index; CAP = controlled attenuation parameter; NAFLD/NASH = Nonalcoholic fatty liver disease / nonalcoholic steatohepatitis

	All patients (n=2346)	S0 (n=750)	S1 (n=623)	S2 (n=501)	S3 (n=472)
Female, n (%)	1147 (48.9%)	346 (46.1%)	306 (49.1%)	251 (50.1%)	244 (51.7%)
Age, yr, mean \pm SD	46.5 \pm 14.5	40.6 \pm 15.4	49.1 \pm 14.0	51.4 \pm 12.5	47.3 \pm 12.6
BMI ¹ , kg/m ² , mean \pm SD	31.4 \pm 9.1	26.2 \pm 7.7	32.5 \pm 8.9	33.4 \pm 8.2	36.3 \pm 8.4
< 25	673 (29.5%)	425 (56.7%)	127 (20.4%)	75 (15.0%)	46 (9.7%)
\geq 25 to < 30	530 (23.2%)	168 (22.4%)	158 (25.4%)	127 (25.3%)	77 (16.3%)
\geq 30 to < 35	322 (14.1%)	41 (5.5%)	98 (15.7%)	102 (20.4%)	81 (17.2%)
\geq 35 to < 40	304 (13.3%)	39 (5.2%)	75 (12.0%)	77 (15.4%)	113 (23.9%)
\geq 40 to < 45	265 (11.6%)	35 (4.7%)	80 (12.8%)	59 (11.8%)	91 (19.3%)
\geq 45	189 (8.3%)	24 (3.2%)	55 (8.8%)	49 (9.8%)	61 (12.9%)
Aetiology, n(%)					
NAFLD/NASH	1277 (54.4%)	96 (12.8%)	376 (60.4%)	394 (78.6%)	411 (87.1%)
Hepatitis B	258 (11.0%)	195 (26.0%)	44 (7.1%)	11 (2.2%)	8 (1.7%)
Hepatitis C	216 (9.2%)	130 (17.3%)	54 (8.7%)	21 (4.2%)	11 (2.3%)
ALD	285 (12.1%)	76 (10.1%)	102 (16.4%)	69 (13.8%)	38 (8.1%)
Other	310 (13.2%)	253 (33.7%)	47 (7.5%)	6 (1.2%)	4 (0.8%)
Diabetes mellitus type 2 ² , n (%)					
NAFLD/NASH	513 (40.9%)	9 (9.4%)	156 (41.5%)	188 (47.7%)	160 (38.9%)
Hepatitis B	16 (6.3%)	8 (4.1%)	2 (4.5%)	4 (36.4%)	2 (25.0%)
Hepatitis C	18 (8.5%)	9 (6.9%)	5 (9.3%)	3 (14.3%)	1 (9.1%)
ALD	37 (13.6%)	6 (7.9%)	14 (13.7%)	11 (15.9%)	6 (15.8%)
Other	48 (17.3%)	41 (16.2%)	5 (10.6%)	2 (33.3%)	0 (0.0%)
ALT ³ , U/L, median [IQR]					
Female	36 [23, 62]	33 [23, 53]	30 [23, 54]	35 [23, 77]	47 [23, 62]
Male	51 [33, 89]	48 [33, 67]	43 [33, 83]	53 [33, 110]	71 [33, 89]
AST ⁴ , U/L, median [IQR]					
Female	33 [23, 53]	30 [23, 48]	30 [23, 42]	30 [23, 70]	42 [23, 53]
Male	40 [29, 57]	39 [29, 55]	36 [29, 61]	40 [29, 70]	51 [29, 57]
Time between CAP and biopsy, n (%)					
CAP > 14 days before biopsy	20 (0.9%)	1 (0.1%)	3 (0.5%)	6 (1.2%)	10 (2.1%)
CAP 2–14 days before biopsy	457 (19.5%)	70 (9.3%)	116 (18.6%)	98 (19.6%)	173 (36.7%)
CAP within 1 day of biopsy	1714 (73.1%)	667 (88.9%)	433 (69.5%)	345 (68.9%)	269 (57%)
CAP 2–14 days after biopsy	96 (4.1%)	9 (1.2%)	49 (7.9%)	29 (5.8%)	9 (1.9%)
CAP > 14 days after biopsy	59 (2.5%)	3 (0.4%)	22 (3.5%)	23 (4.6%)	11 (2.3%)
Liver stiffness, kPa, median [IQR]					
M-probe ⁵	7.8 [5.7, 9.6]	6.6 [5.7, 12.0]	8.2 [5.7, 13.1]	8.8 [5.7, 13.5]	9.4 [5.7, 9.6]
XL-probe ⁶	6.9 [5.3, 8.8]	6.3 [5.3, 9.9]	6.6 [5.3, 10.7]	7.3 [5.3, 12.4]	7.9 [5.3, 8.8]
Fibrosis staging ⁷ , n (%)					
F0	528 (23.1%)	264 (35.2%)	130 (20.9%)	74 (14.8%)	60 (12.7%)
F1	790 (34.6%)	208 (27.7%)	221 (35.5%)	182 (36.3%)	179 (37.9%)
F2	423 (18.5%)	106 (14.1%)	113 (18.1%)	98 (19.6%)	106 (22.5%)
F3	316 (13.8%)	66 (8.8%)	75 (12.0%)	89 (17.8%)	86 (18.2%)
F4	225 (9.9%)	78 (10.4%)	65 (10.4%)	52 (10.4%)	30 (6.4%)

¹ Data available from 2283 patients

² Data available from 2272 patients

³ Data available from 2307 patients

⁴ Data available from 2282 patients

⁵ Data available from 1478 patients

⁶ Data available from 1270 patients

⁷ Data available from 2282 patients

Diagnostic Performance of CAP for Assessing Steatosis

The results of ROC analyses are provided in **Table 2** by aetiology and for the XL probe alone in patients with NAFLD. Only data from the correct probe according to the definition in the Methodology section were included. For the XL probe in patients with NAFLD, the optimal cut-off to distinguish S0-1 from S2-3 was 317 dB/m (95% CI 306 to 334) and only marginally higher compared to the value of 310 dB/m, found in the total cohort of patients with NAFLD for which the XL probe was used in 73% of patients. Notably it was substantially higher than for patients with viral hepatitis, ALD or other aetiologies. Note that all but 15 of the patients with ALD came from a single study,⁴¹ but that the Youden optimization chosen in **Table 2** complements other choices in that paper, which was optimized to have sensitivity/specificity of 0.9. The optimum cut-off in patients with ALD for S0 vs S1-3 was paradoxically higher than for S0-1 vs S2-3 (274 vs 268 dB/m), but the confidence intervals were wide (over 50 dB/m) and the shifts in sensitivity and specificity suggests that their sum remained fairly constant over a large range of CAP values. The large differences between aetiologies demonstrate that they must be considered separately in many analyses. For some analyses, we choose to focus on NAFLD, which is where the XL probe is most relevant, cf. the proportion of XL probe use according to aetiology in **Table 2**.

Supplementary table 2 shows the diagnostic performance in NAFLD patients undergoing bariatric surgery vs non-bariatric patients and in those with BMI above and below 30 kg/m². The distribution of steatosis grades depends on cohort type and diagnostic performance tends to be better in the bariatric cohorts, but relevant differences vanish when distinguishing by BMI, though optimal cut-offs may be affected. The correct probe choice in the NAFLD cohort is a matter of some debate, and moreover the penetration depth of even the XL probe can be a limiting factor in morbidly obese patients.^{42, 13} In **Supplementary tables 3 and 4** we provide the diagnostic performance using the manufacturer's recommendation for probe choice and find analogous results, e.g. for the NAFLD cohort, optimal cut-offs are within one dB/m.

Table 2: Diagnostic properties of CAP for assessing histologically defined steatosis grade as estimates (95% confidence intervals). Only data from the correct probe choice were included. If the available data were too scant, then estimates were not provided, e.g. only 4 patients with "other" aetiologies had S3. AUC = Area under the curve; CAP = Controlled attenuation parameter; HBV = Hepatitis B virus; HCV = Hepatitis C virus; NAFLD = Non-alcoholic fatty liver disease.

	S0 vs S1-3	S0-1 vs S2-3	S0-2 vs S3
NAFLD XL probe (n=930, 95% BMI > 30 kg/m ²)			
AUC	0.819 (0.769 to 0.869)	0.754 (0.720 to 0.787)	0.717 (0.684 to 0.751)
Prevalence	0.911	0.630	0.346
Optimal cut-off (dB/m)	297 (287 to 323)	317 (306 to 334)	333 (320 to 340)
Sensitivity	0.798 (0.771 to 0.824)	0.783 (0.748 to 0.816)	0.761 (0.714 to 0.808)

Specificity	0.735 (0.639 to 0.831)	0.628 (0.574 to 0.679)	0.597 (0.559 to 0.636)
NAFLD (n=1274, 73% XL, 76% BMI > 30 kg/m ²)			
AUC	0.807 (0.758 to 0.855)	0.736 (0.707 to 0.765)	0.711 (0.682 to 0.740)
Prevalence	0.925	0.630	0.322
Optimal cut-off (dB/m)	294 (286 to 313)	310 (305 to 321)	331 (319 to 340)
Sensitivity	0.790 (0.767 to 0.813)	0.790 (0.761 to 0.817)	0.718 (0.674 to 0.762)
Specificity	0.740 (0.646 to 0.823)	0.589 (0.557 to 0.634)	0.621 (0.589 to 0.652)
HBV or HCV (n=472, 10% XL, 8% BMI > 30 kg/m ²)			
AUC	0.769 (0.724 to 0.814)	0.847 (0.794 to 0.900)	—
Prevalence	0.311	0.106	0.040
Optimal cut-off (dB/m)	230 (209 to 266)	264 (238 to 285)	—
Sensitivity	0.714 (0.640 to 0.789)	0.760 (0.640 to 0.880)	—
Specificity	0.680 (0.628 to 0.729)	0.794 (0.754 to 0.832)	—
ALD (n=284: 8% XL, 20% BMI > 30 kg/m ²)			
AUC	0.765 (0.704 to 0.826)	0.766 (0.711 to 0.821)	0.802 (0.733 to 0.871)
Prevalence	0.732	0.377	0.134
Optimal cut-off (dB/m)	274 (236 to 291)	268 (263 to 310)	291 (285 to 319)
Sensitivity	0.644 (0.577 to 0.707)	0.860 (0.794 to 0.925)	0.868 (0.763 to 0.974)
Specificity	0.790 (0.697 to 0.882)	0.554 (0.480 to 0.627)	0.679 (0.622 to 0.736)
Other (n=309, 7% XL, 16% BMI > 30 kg/m ²)			
AUC	0.687 (0.612 to 0.762)	—	—
Prevalence	0.184	0.032	0.013
Optimal cut-off (dB/m)	244 (205 to 262)	—	—
Sensitivity	0.684 (0.561 to 0.807)	—	—
Specificity	0.659 (0.599 to 0.718)	—	—

If instead of the Youden cut-off approach, the sensitivity is set at 0.9, then the optimum in NAFLD patients is 267, 286 and 297 dB/m for S0 vs S1-3, S0-1 vs S2-3 and S0-2 vs S3 respectively. The values for specificity at these optima are 0.500, 0.394 and 0.344 respectively. If the specificity is set at 0.9, optimal cut-off values are 331, 372 and 385 dB/m for S0 vs S1-3, S0-1 vs S2-3 and S0-2 vs S3 respectively with commensurate sensitivities of 0.520, 0.255 and 0.224.

The diagnostic performance in patients with NAFLD using two sets of established cut-offs with and without adjustment of CAP values for aetiology, diabetes and BMI according to a previous suggestion derived from M-probe data can be found in **Supplementary table 5**.^{11, 14} The cut-offs taken from an IPDMA with mixed aetiology and the M-probe result in sensitivities above 0.9, but very poor specificities ranging from 0.25 to 0.42. Even NAFLD derived cut-offs do not lead to satisfactory sensitivity and specificity, the range being from 0.57 to 0.74. Adjusting CAP values, but retaining the aforementioned IPDMA cut-offs, leads to sensitivity and specificity roughly comparable to those found with NAFLD derived cut-offs.

Using ordinal regression, it is possible to estimate the probability for steatosis grade at a given CAP value, **Figure 2**. Such probabilities are closely related to PPV and NPV and depend strongly on prevalence. In some

prevalence scenarios taken from the papers in this analysis, there are CAP values for which three or even all four steatosis grades are probable and distinction between them is therefore poor.

355 *Diagnostic performance across studies*

Although one works with individual patient data in an IPDMA, it cannot be treated as a single study. To assess potential study effects, we present areas under the curve, which are independent of cut-offs, between the original sources of our data and the pooled result from this IPDMA, **Figure 3**. The figures were recalculated using only the patients included in this IPDMA and a standardized optimization Youden method. There is substantial
360 variation between studies and increasing NAFLD prevalence relative to the other liver diseases tends to lead to poorer diagnostic performance.

Probe Choice and Relation to Aetiology

The device offers two probes (M and XL), which can either be selected manually or by an automated probe selection software implemented in the most recent version of the device. The study designs considered here
365 foresaw automated choice of probes for 895 patients (4 studies), use of both probes for 511 patients (5 studies), based on BMI for 443 patients (1 study), based on SLD for 421 patients (2 studies) and only used the XL probe for 76 patients (1 study). However, in some cases, if the selected probe failed, then the other one was used. The XL probe was correctly chosen according to our definitions in 1050 patients and the M probe in 1289 patients including 20 cases with BMI > 35 kg/m² but appropriate SLD. For seven patients, the correct probe choice could
370 not be determined since BMI was unavailable.

Note that essentially all correct XL measurements were made on NAFLD patients (930/1050, 89%). In contrast to the previous IPDMA,¹¹ steatosis grade S0 was dominated by aetiologies other than NAFLD (654/750, 87%) whereas all other aetiologies were underrepresented for S1-S3 (415/1596, 26%), **Supplementary figure 2**.

Notably, 75% of the NAFLD cohort with S0 (72/96 patients) were bariatric surgery patients.

375

Covariates affecting CAP including the choice of M or XL probe

The M and XL probes are designed for different anthropometric phenotypes based on skin-to-liver-capsule distance and therefore a direct comparison of their properties is clinically relevant in patients for whom either probe is conceivable. Moreover, such a comparison in all patients sheds light on the diagnostic and technical
380 differences between the probes. CAP values from both probes were available in 527 patients (58% NAFLD). A

Bland-Altman plot in **Supplementary figure 3** shows that there is little bias (4.4 dB/m) and essentially no dependence on the mean CAP value over a wide range of BMI values. However, the typical discrepancy between the M and XL probe in a given patient is large (the mean absolute difference is 30 dB/m and a 95% predictive interval ranges from -78 to 82 dB/m at a mean CAP value of 250 dB/m).

385 For all patients, the correct probe (M or XL) was also included as a covariate in a linear mixed model in which study was the random term. The results of this multivariate analysis are provided in **Table 3** and show that CAP values for patients with steatosis differ by about 30, 62 and 81 dB/M from those without for S1, S2 and S3, respectively. Aetiology, BMI, sex, aspartate transaminase (AST) and diabetes also play a role, but the effect of probe choice is small. The standard deviation of the random term is 10.3 dB/m with a residual of 46.5 dB/m
 390 and an ANOVA indicates that it differs significantly from a linear model without the random term ($p < 0.001$), indicating that differences between studies are not negligible despite inclusion of covariates. However, estimates do not change by much in a linear model with study as a fixed effect despite accounting for steatosis grades, and although the term study is significant as a whole ($p < 0.001$), no single study differs significantly from the arbitrarily chosen reference study. Similar results are found in patients with NAFLD alone, except that AST is no
 395 longer identified as a relevant or significant covariate (**Supplementary table 6**). If one adjusts the CAP value in NAFLD patients by 10 dB/m for diabetics and males and by 2.6 dB/m per BMI point above/below 25 kg/m² up to a maximum of 50 kg/m², then AUC changes from 0.807 to 0.826 (S0 vs S1-3), 0.736 to 0.762 (S0-1 vs S2-3) and 0.711 to 0.702 (S0-2 vs S3).

400 **Table 3:** Estimates and confidence intervals are shown from the fixed terms in a mixed model with CAP value as the dependent variable and study as a random term. The estimates describe how each variable is associated with the CAP result, even after accounting for the others. For example, patients with diabetes will have CAP values an average of 13.6 dB/m higher, even after accounting for aetiology steatosis grade, BMI, sex, ALT and AST. For a patient with a BMI of 27 kg/m², mean CAP will be $2 \times 2.57 = 5.14$ dB/m greater than for a patient with a BMI of 25 kg/m². ALD = alcoholic liver disease; ALT = Alanine transaminase; AST = Aspartate
 405 transaminase; BMI = Body mass index; CAP = Controlled attenuation parameter; CI = Confidence interval; HBV = Hepatitis B virus; HCV = Hepatitis C virus; NALFD/NASH = Non-alcoholic fatty liver disease/ Non-alcoholic steatohepatitis.

	Estimate (dB/m)	95% CI (dB/m)
Intercept	239	219 to 259
Steatosis (S0 is reference)		
S1	30.4	24.2 to 36.8
S2	62.1	55.1 to 69.4
S3	81.0	73.6 to 88.5
Aetiology (ALD as reference)		
HBV	-17.2	-34.8 to 0.2
HCV	-18.2	-35.6 to -0.8
NAFLD/NASH	0.6	-16.2 to 17.4
Other	-20.6	-37.7 to -3.5
XL compared to M probe	6.5	-0.5 to 12.6

BMI (for each kg/m ² increase compared to 25 kg/m ²)	2.57	2.11 to 2.97
Diabetes mellitus type 2	13.6	8.7 to 18.8
Sex (male vs female)	12.0	7.7 to 16.4
ALT (per doubling of value)	1.06	-2.07 to 4.34
AST (per doubling of value)	-3.81	-7.46 to -0.29

410 For some patients, the M and XL probes may both be viable alternative choices, i.e. when penetration depth is close to the cross-over from the M to the XL probe. Unfortunately, SLD was unavailable in data sets for which both probes were used. Hence, we consider NAFLD patients with $23 < \text{BMI} < 30 \text{ kg/m}^2$, in which more than half (56%) had a discrepancy, when steatosis categories are allocated using the cut-offs from **Table 2**. There is a difference of at least two categories in 27/135 (20%) cases, **Supplementary table 7**. The M-probe yielded the correct result in 33/135 (24%) and the XL probe in 37/135 cases (27%).

Reliability criteria in NAFLD patients

Diagnostic properties could be affected by unreliable measurements. If the threshold $\text{IQR} > 40 \text{ dB/m}$ is used, then 32% of measurements are defined to be unreliable. However, the correlation between CAP and percentage hepatocytes does not improve, changing only from 0.452 in the former to 0.449 in the latter case, see **Supplementary figure 3a**. If instead the threshold is based on IQR/median , then 3%, 11% and 46% are deemed unreliable for cut-off of 0.3, 0.2 and 0.1 respectively. In fact, the analogous correlations become worse as the “reliability” criterion is made more stringent ($r=0.436, 0.399$ and 0.337 respectively, see **Supplementary figure 3b**). Hence, we were not able to find evidence that any cut-off value provided value in improving performance.

425

DISCUSSION

This comprehensive IPDMA on the diagnostic capabilities of CAP including the XL probe for non-invasive grading of hepatic fat found that diagnostic properties depend substantially on aetiology. Higher grades of steatosis in patients with NAFLD could not be distinguished reliably for individual patients and the areas under the curve for the XL probe were 0.82, 0.75 and 0.72 with optimal cut-offs at 297, 317 and 333 dB/m for >S0, >S1 and >S2 respectively. Mean CAP values from the available M and XL double measurements in a given patient were similar although the variation between them was high, resulting in misclassification of 50% of NAFLD cases (**Supplementary figure 3**). Diagnostic accuracy did not improve by applying proposed reliability criteria.^{36, 37} After adjusting for steatosis grade, we verified that CAP depended significantly on aetiology, diabetes and BMI and found moreover that CAP for males is about 13 dB/m higher than for females. Unexpectedly, diagnostic performance in bariatric surgery patients was somewhat better, which may in part be due to highly standardized recruitment and procedures as well as large biopsy samples.

Although our quantitative results agree on the whole with the published literature (cf. Figure 4), the picture that emerges when the results are viewed in this pooled fashion is new, and the conclusions we draw are different. Individual studies suggest that the diagnosis of steatosis in NAFLD with CAP works well with the XL probe in obese patients, i.e. that the AUC for S0 vs S1-3 is high. However, some of these studies contain cohorts with mixed aetiologies in which S0 is dominated by non-NAFLD patients.^{21, 16, 15} Others are pure NAFLD cohorts, but with a small number of S0 patients.^{26, 14, 23, 13, 18} The definition of NAFLD formally excludes S0 so that their occurrence merits further consideration. In our data, S0 patients come either from studies with bariatric surgery cohorts^{18, 23, 13} where histology is routinely obtained during surgery and irrespective of liver disease severity, or they come from a screening scenario where the majority of S0 patients were ultimately classified as non-NAFLD.¹⁴ Notably, the S0 NAFLD cases did not have advanced fibrosis suggestive of burnout NASH. Because the S0 group is small and atypical, the important clinical question: “can steatosis be detected by CAP in a screening setting for subjects at risk for NAFLD” cannot be answered adequately with currently available data.

Our group conducted a previous IPDMA on M probe CAP and found good results, with AUCs ranging from 0.82 to 0.88.¹¹ The data were restricted to “leaner” subjects (BMI > 35 kg/m² was an exclusion criterion) with M probe measurements and were dominated by viral hepatitis. Moreover, there were almost no S0 patients classified as having NAFLD and diagnostic properties for this aetiology alone could not be provided. An early conventional meta-analysis did treat diagnostic properties in NAFLD, but looked mainly at M probe data despite

455 many obese subjects and concluded that diagnostic capabilities should be viewed with caution.⁴³ Only two of the
nine studies considered there used the XL probe and thus qualified for the current analysis and only one could be
included. Our current data also suggest the need to be cautious with CAP for steatosis grading in NAFLD even if
the XL probe is used, with AUCs of 0.75 (S0-1 vs S2-3) and 0.72 (S0-2 vs S3), cf. Table 2.

In the current meta-analysis, patients with viral hepatitis tended to have higher BMIs compared to our original
460 IPDMA¹¹ because of the availability of the XL probe. The low prevalence of high-grade steatosis and relatively
small total numbers meant that diagnostic performance could not be analysed in great depth for this aetiology.
AUC was slightly lower here for S0 vs S1-3 compared to the previous IPDMA (0.77 vs 0.82), whereas it was
comparable for S0-1 vs S2-3 (0.85 vs 0.86). Patients with ALD in the current paper had a fairly high prevalence
of obesity (20%) despite infrequent use of the XL probe (8%), but they derived almost exclusively from a single
465 study,²⁵ suggesting that more data in this aetiology are needed.

There has been some debate about optimal cut-offs and how they may depend on aetiology, probe (M vs XL)
and anthropometry.^{26, 14, 11, 44, 45} Considering the different penetration depths of the probes, it is essential to
distinguish between two types of comparison: (i) between two probes used in a given patient where such data are
available, in which case small mean differences were found of less than 10 dB/m^{26, 44, 46} and (ii) between probes
470 each used in the appropriate cohort (majority of data), in which case large differences in cut-offs are found
unless models correct for anthropometry and co-morbidities. Here we can confirm both observations but
conclude that the small differences between the probes can be neglected if the correct probe is used. The choice
of cut-off then may benefit from corrections for aetiology, diabetes and BMI as recommended in,¹¹ which for
NAFLD in the current paper would lead to a shift from e.g. 280 dB/m (S0-2 vs S3) to about 330 dB/m, which is
475 close to 337 dB/m as found by Eddowes et al.¹⁴ Without such corrections, sensitivity is very high, but specificity
is extremely low. Other authors have already noted that corrections do not necessarily improve AUC or other
markers of overall diagnostic performance.^{26, 37} Corrections do however lead to a shift toward more equal
balancing between sensitivity and specificity that could be relevant if both ruling in and ruling out steatosis are
of interest. When only one of the two is the diagnostic goal, then it is typical to optimize based on either
480 sensitivity or specificity of say 0.9, see e.g. Eddowes 2019.¹⁴ For a specificity of 0.9, this can lead to extremely
high cut-offs, which are no longer of clinical use. In the context of NAFLD, the current analysis suggests that
adjustments may have the potential to improve AUC slightly, but this requires validation.

It is interesting to speculate as to what physiological mechanism could affect CAP measurements in diabetic patients after taking BMI, sex and other covariates into account. There is evidence that microvesicular steatosis is related to progression of NAFLD⁴⁷ meaning that it may be more prevalent among diabetes patients.⁴⁸ Since small fat droplets lead to more signal attenuation, this may lead to higher CAP values. Furthermore, diabetes patients may have a greater prevalence of steatohepatitis, which could also affect CAP. Such considerations could also be relevant in explaining aetiology specific differences. Beyond physiological mechanisms, the very different prevalence of steatosis depending on aetiology leads to different optimal cut-offs in ROC analyses.

The non-invasive quantification of hepatic fat requires critical appraisal since its clinical implications are a matter of ongoing debate.⁴⁹ Although fat accumulation in hepatocytes is considered a key mechanism in the natural history of fatty liver disease, long-term observational studies have not found that the histological degree of steatosis correlates with clinical endpoints, e.g. Angulo et al.⁵⁰ This finding challenges the need for estimating hepatic fat, but associations have been found between liver fat and progression of fibrosis⁵¹ and between liver fat and metabolic syndrome, even after adjusting for NASH.⁵² Moreover, current outcome data are biased due to the need for an invasive reference standard and fairly small sample sizes compared to blood pressure in cardiology, for example. Furthermore, liver fat assessed by histology refers to a proportion of affected cells by surface area, whereas MRI and CAP consider other conceptual notions that still require appropriate long-term observation of liver fat.⁹ In addition, the value of serial assessment of liver fat quantification remains to be defined for prognosis and treatment efficacy of intervention.⁷

Another potential clinical benefit for CAP can be in identifying the small proportion of NAFLD candidates with S0 who ultimately turn out to have alternative diagnoses as shown in Eddowes et al.¹⁴ Studies focusing on screening cohorts cannot use liver biopsies as a reference standard, but could use MRI, which is very sensitive to hepatic fat.⁸ Up to now only a few studies have compared both MRI and CAP with histology with a modest sample size^{20, 53, 54} or without S0 patients,⁵⁵ but all of them concluded that MRI outperforms CAP. Such studies can also address the value of continuous tests for steatosis severity, which is not possible with traditional steatosis grading. However, the clinical availability and costs of MRI, imply that combinations of screening tools need to be considered.⁹

The data presented here suggest that CAP could well be of epidemiological value but raise questions about what role it could play in the GP's toolbox. Previous analyses have shown good diagnostic performance in hepatitis patients using the M probe,¹¹ and the current analysis suggests only a slight deterioration when extending the

population to patients with higher BMI and use of the XL probe. However, it is of limited diagnostic value for grading steatosis in obese NAFLD patients, which emphasizes inherent technical limitations of this ultrasound-based steatosis quantification. Using CAP, ultrasound signal attenuation is not exclusively induced by
515 hepatocellular lipid droplets but is also related to subcutaneous tissue and abdominal wall⁵⁶ properties which become especially relevant in patients with metabolic risk factors and may explain our observations.

Despite such considerations, CAP still represents the first highly standardized and evaluated approach using ultrasound signal modulation for steatosis quantification. Technical modifications of CAP such as special algorithms for extremely obese patients¹³ could improve diagnostic precision. In addition, further developments
520 by other manufacturers utilize two-dimensional B-mode image control of the target region, which enhances performance in pilot studies and merit further evaluation.^{57, 58}

Strengths of our analysis are that it relies on the largest comprehensive database on biopsy-controlled CAP measurements and that the mean time interval between the index (CAP) and reference tests (biopsy) is very short. The data derive from many expert international groups with only small differences between studies. The
525 large number of paired XL and M probe measurements in overweight patients without *a priori* indication for one or the other facilitated important comparisons between the two. This pooling of the available NAFLD patients exposed the fact that a considerable proportion of this population derives from bariatric surgery, but not from a typical screening target population and should be complemented in new studies.

A general weakness of studies with invasive reference standards is the so-called “verification bias”, see e.g.
530 O’Sullivan et al.⁵⁹ It is incorrect to believe that sensitivity and specificity estimated from a complete case analysis are unbiased compared to a population without (indication for) biopsy, which again underscores the need for CAP screening studies with MRI as the reference standard. The high prevalence of significant fibrosis illustrates that indication for biopsy was based on the suspicion of advanced liver disease in many of the patients. Moreover, bariatric surgery patients are recruited with less emphasis on liver disease and tended to have fewer
535 “grey zone” patients with S1 and S2 so that their diagnostic performance was better. This may also hold for screening populations of interest without indication for liver biopsy. A related issue is the spectrum effect, describing a dependence on sensitivity and specificity with regard to patient characteristics.⁶⁰ This can induce spectrum bias if populations are selected based on such characteristics and results are generalized to a different target population and this may explain in part the non-negligible difference between studies. The QUADAS 2
540 assessment showed that there is some potential for bias as a result of patient selection and precise use of the

index test. These are expected to be small compared to the verification and spectrum biases, however, when CAP is used in a screening population. A specific limitation of this IPDMA is that we could not include data from the US NASH Clinical Research Network because of rigid federal NIH restrictions.^{20, 17, 27} However, these studies in total are expected to have supplied data from about 300 patients (Supplementary table 1) and only about 20 with S0. Given the somewhat worse diagnostic performance in Table 4 of Siddiqui 2018, this lack of data can be expected to bias our results slightly toward optimistic estimates. It is not feasible in an intercontinental diagnostic IPDMA to obtain central histological readings. However, since steatosis is known to have good inter- and intra-observer agreement, in contrast to other NASH parameters,^{61, 32} this is likely only to have led to a small bias, if any. The choice of correct probe in the studies analysed here did not always follow the manufacturer's recommendations and the effect of this on performance would be interesting, but was not possible in this IPDMA given the nature of the data. Future studies should firmly adhere to the recommendations. Finally, the strict definition of NAFLD precluding concomitant diseases could be inherently problematic since it neglects metabolically driven processes in liver disease of other aetiology. A recent proposal for a revision of the definition of metabolic liver disease addresses this issue, but was not available in the current data set.⁶² Such considerations could also be relevant for patients with viral hepatitis and concomitant fatty liver disease.

This analysis completes a decade of CAP research for the grading of steatosis, but does not focus on recent advancements for the detection of steatohepatitis, which combine CAP with LSM and AST and may be useful for guiding medical intervention in the future.⁶³ However, the imprecision of CAP-based steatosis grading, especially when important covariates are left unconsidered, sheds light on the capabilities of non-invasive NASH detection. This highlights the necessity for several independent studies with a variety of designs.

In conclusion, CAP cut-offs vary according to aetiology, and effectively recognize significant steatosis in patients with viral hepatitis. CAP values alone relative to rigid cut-offs provide little clinically valuable information for grading steatosis in patients with NAFLD despite use of the XL probe. Based on a subset of the data, the difference between M and XL probe values in a given patient may be large, but their mean difference in a given population is quite small. Current data on S0 in the context of NAFLD are scant and heavily influenced by bariatric surgery, so that firm conclusions on diagnostic performance cannot yet be drawn. Knowledge of prevalence and covariates may help interpret CAP use and the application in screening scenarios requires further research beyond the traditional histological reference standard.

570

Conflict of interest

DP, TK and JW received unrestricted research grants from Echosens, Paris France and TK participated in a clinical advisory board meeting. MS was an Echosens scientific employee in the R&D department, but contributed in her role as scientist and author of papers in this meta-analysis. Echosens provided the FibroScan® device to Antoine-Béclère hospital, but the authors CSV, GP and SN did not receive funding from the manufacturers for carry out their research. VW served as a consultant and speaker for Echosens. No other authors have relevant conflicts of interest.

Detailed Funding including original data sources

This project is supported in part by the German Federal Ministry of Education and Research (BMBF, FKZ 01EO1001). The company Echosens provided funding, but did not influence study design and was not involved in data analysis. Data collection at Odense University Hospital received support from Innovation Fund Denmark, the Challenge Grant “MicrobLiver” grant number NNF15OC0016692 from the Novo Nordisk Foundation and the European Union’s Horizon 2020 research and innovation programme for “GALAXY” grant agreement number 668031. Data collection at the Federal University of Rio de Janeiro received support from the National Council for Research and Technology – CNPq, Brazil. QMA is a Newcastle NIHR Biomedical Research Centre investigator. PNN was supported by the National Institute of Health Research (NIHR) Birmingham Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Author Contributions

DP, VB, JW, TK: study concept and design, literature search, collected data from participating centres, data review, verification and analysis, interpretation of data, drafting of the manuscript, statistical analysis, obtained funding

Remaining authors: drafting of protocol, obtained original data, coded and provided data, interpretation of data, critical revision of the manuscript

REFERENCES

1. Castera L, Friedrich-Rust M, Loomba R. Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2019; 156(5):1264-1281.e4.
- 600 2. European Association for Study of Liver, Asociacion Latinoamericana para el Estudio del Hgado. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *Journal of Hepatology* 2015; 63(1):237–64.
3. Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwithaya P et al. Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2015; 149(2):389-97.e10.
- 605 4. **Ekstedt M, Hagström H**, Nasr P, Fredrikson M, Stål P, Kechagias S et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology* 2015; 61(5):1547–54.
5. McPherson S, Hardy T, Henderson E, Burt AD, Day CP, Anstee QM. Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. *Journal of Hepatology* 2015; 62(5):1148–55.
- 610 6. Liu K, Wong VW-S, Lau K, Du Liu S, Tse Y-K, Yip TC-F et al. Prognostic Value of Controlled Attenuation Parameter by Transient Elastography. *Am J Gastroenterol* 2017 [cited 2017 Nov 7]; 112(12):1812–23.
7. Wong VW-S. Predicting NASH response with liver fat: Are we back to square one? *Journal of Hepatology* 2020; 72(3):386–8.
- 615 8. Caussy C, Reeder SB, Sirlin CB, Loomba R. Noninvasive, Quantitative Assessment of Liver Fat by MRI-PDF as an Endpoint in NASH Trials. *Hepatology* 2018 [cited 2020 Mar 8]; 68(2):763–72.
9. Karlas T, Petroff D, Wiegand J. Collaboration, Not Competition: The Role of Magnetic Resonance, Transient Elastography, and Liver Biopsy in the Diagnosis of Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2017; 152(3):479–81.
- 620 10. Sasso M, Beaugrand M, Ledinghen V de, Douvin C, Marcellin P, Poupon R et al. Controlled Attenuation Parameter (CAP): A Novel VCTE™ Guided Ultrasonic Attenuation Measurement for the Evaluation of Hepatic Steatosis: Preliminary Study and Validation in a Cohort of Patients with Chronic Liver Disease from Various Causes. *Ultrasound in Medicine & Biology* 2010; 36(11):1825–35.
- 625 11. **Karlas T, Petroff D**, Sasso M, Fan J-G, Mi Y-Q, Lédighen V de et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. *Journal of Hepatology* 2017; 66(5):1022–30.
12. Sasso M, Audiere S, Kengang A, Gaouar F, Corpechot C, Chazouilleres O et al. Liver Steatosis Assessed by Controlled Attenuation Parameter (CAP) Measured with the XL Probe of the FibroScan: A Pilot Study Assessing Diagnostic Accuracy. *Ultrasound in Medicine & Biology* 2016; 42(1):92–103.
- 630 13. Somda S, Lebrun A, Tranchart H, Lamouri K, Prevot S, Njike-Nakseu M et al. Adaptation of controlled attenuation parameter (CAP) measurement depth in morbidly obese patients addressed for bariatric surgery. *PLoS ONE* 2019 [cited 2020 Mar 6]; 14(5):e0217093.
14. Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D et al. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2019; 156(6):1717–30.
- 635 15. Baumeler S, Jochum W, Neuweiler J, Bergamin I, Semela D. Controlled attenuation parameter for the assessment of liver steatosis in comparison with liver histology: A single-centre real life experience. *Swiss Med Wkly* 2019; 149:w20077.
- 640 16. Rout G, Kedia S, Nayak B, Yadav R, Das P, Acharya SK et al. Controlled Attenuation Parameter for Assessment of Hepatic Steatosis in Indian Patients. *J Clin Exp Hepatol* 2019; 9(1):13–21.
17. Vuppalanchi R, Siddiqui MS, van Natta ML, Hallinan E, Brandman D, Kowdley K et al. Performance characteristics of vibration-controlled transient elastography for evaluation of nonalcoholic fatty liver disease. *Hepatology* 2018 [cited 2018 Apr 9]; 67(1):134–44.
- 645 18. Garg H, Aggarwal S, Shalimar, Yadav R, Datta Gupta S, Agarwal L et al. Utility of transient elastography (fibroscan) and impact of bariatric surgery on nonalcoholic fatty liver disease (NAFLD) in morbidly obese patients. *Surg Obes Relat Dis* 2018; 14(1):81–91.
19. Lai L-L, Wan Yusoff WNI, Vethakkan SR, Nik Mustapha NR, Mahadeva S, Chan W-K. Screening for non-alcoholic fatty liver disease in patients with type 2 diabetes mellitus using transient elastography. *J Gastroenterol Hepatol* 2019; 34(8):1396–403.
- 650

20. Park CC, Nguyen P, Hernandez C, Bettencourt R, Ramirez K, Fortney L et al. Magnetic Resonance Elastography vs Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients With Biopsy-Proven Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2017; 152(3):598-607.e2.
- 655 21. Lédighen V de, Hiriart J-B, Vergniol J, Merrouche W, Bedossa P, Paradis V. Controlled Attenuation Parameter (CAP) with the XL Probe of the Fibroscan®: A Comparative Study with the M Probe and Liver Biopsy. *Dig Dis Sci* 2017;1–9. Available from: URL: <https://link.springer.com/content/pdf/10.1007%2Fs10620-017-4638-3.pdf>.
- 660 22. Chan W-K, Nik Mustapha NR, Wong GL-H, Wong VW-S, Mahadeva S. Controlled attenuation parameter using the FibroScan® XL probe for quantification of hepatic steatosis for non-alcoholic fatty liver disease in an Asian population. *United European Gastroenterol J* 2017 [cited 2020 Mar 6]; 5(1):76–85.
23. Naveau S, Voican CS, Lebrun A, Gaillard M, Lamouri K, Njiké-Nakseu M et al. Controlled attenuation parameter for diagnosing steatosis in bariatric surgery candidates with suspected nonalcoholic fatty liver disease. *Eur J Gastroenterol Hepatol* 2017; 29(9):1022–30.
- 665 24. Cardoso AC, Cravo C, Calçado FL, Rezende G, Campos CFF, Neto JMA et al. The performance of M and XL probes of FibroScan for the diagnosis of steatosis and fibrosis on a Brazilian nonalcoholic fatty liver disease cohort. *Eur J Gastroenterol Hepatol* 2020 [cited 2020 Mar 6]; 32(2):231–8.
25. Thiele M, Rausch V, Fluhr G, Kjærgaard M, Piecha F, Mueller J et al. Controlled attenuation parameter and alcoholic hepatic steatosis: Diagnostic accuracy and role of alcohol detoxification. *Journal of Hepatology* 2018 [cited 2019 Apr 2]; 68(5):1025–32.
- 670 26. Chan W-K, Nik Mustapha NR, Mahadeva S, Wong VW-S, Cheng JY-K, Wong GL-H. Can the same controlled attenuation parameter cut-offs be used for M and XL probes for diagnosing hepatic steatosis? *J Gastroenterol Hepatol* 2018 [cited 2019 Apr 4]; 33(10):1787–94.
- 675 27. Siddiqui MS, Vuppalanchi R, van Natta ML, Hallinan E, Kowdley KV, Abdelmalek M et al. Vibration-controlled Transient Elastography to Assess Fibrosis and Steatosis in Patients With Nonalcoholic Fatty Liver Disease. *Clinical Gastroenterology and Hepatology* 2019 [cited 2018 Sep 18]; 17(1):156–63. Available from: URL: [https://www.cghjournal.org/article/S1542-3565\(18\)30405-1/pdf](https://www.cghjournal.org/article/S1542-3565(18)30405-1/pdf).
28. Karlas T, Beer S, Babel J, Busse H, Schaudinn A, Linder N et al. Do we need controlled attenuation parameter adjustment for fibrosis estimation in nonalcoholic fatty liver disease patients? *Hepatology* 2017; 65(6):2126–8. Available from: URL: <http://onlinelibrary.wiley.com/doi/10.1002/hep.29067/full>.
- 680 29. Kleiner DE, Brunt EM, van Natta M, Behling C, Contos MJ, Cummings OW et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005; 41(6):1313–21.
30. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155(8):529–36.
- 685 31. Bedossa P, The French Metavir Cooperative Study Group. Intraobserver and Interobserver Variations in Liver Biopsy Interpretation in Patients with Chronic Hepatitis C. *Hepatology* 1994; 20(1):15–20.
32. Merriman RB, Ferrell LD, Patti MG, Weston SR, Pabst MS, Aouizerat BE et al. Correlation of paired liver biopsies in morbidly obese patients with suspected nonalcoholic fatty liver disease. *Hepatology* 2006 [cited 2020 Mar 12]; 44(4):874–80.
- 690 33. Dietrich CF, Bamber J, Berzigotti A, Bota S, Cantisani V, Castera L et al. EFSUMB Guidelines and Recommendations on the Clinical Use of Liver Ultrasound Elastography, Update 2017 (Long Version). *Ultraschall in der Medizin - European Journal of Ultrasound* 2017; 38(4):e16-e47. Available from: URL: <http://www.thieme-connect.com/products/ejournals/pdf/10.1055/s-0043-103952.pdf>.
- 695 34. **Shen F, Zheng R-D**, Shi J-P, Mi Y-Q, Chen G-F, Hu X et al. Impact of skin capsular distance on the performance of controlled attenuation parameter in patients with chronic liver disease. *Liver Int* 2015; 35(11):2392–400.
35. Boursier J, Zarski J-P, Ledinghen V de, Rousselet M-C, Sturm N, Lebaill B et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology* 2013; 57(3):1182–91.
- 700 36. **Wong VW-S, Petta S**, Hiriart J-B, Cammà C, Wong GL-H, Marra F et al. Validity criteria for the diagnosis of fatty liver by M probe-based controlled attenuation parameter. *Journal of Hepatology* 2017 [cited 2020 Apr 23]; 67(3):577–84.
37. Semmler G, Wöran K, Scheiner B, Unger LW, Paternostro R, Stift J et al. Novel reliability criteria for controlled attenuation parameter assessments for non-invasive evaluation of hepatic steatosis. *United European Gastroenterol J* 2020 [cited 2020 Apr 16]; 8(3):321-331.

- 705 38. López-Ratón M, Rodríguez-Álvarez MX, Suárez CC, Sampedro FG. OptimalCutpoints : An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *J. Stat. Soft.* 2014; 61(8):1–36.
39. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association* 1927 [cited 2020 Sep 17]; 22:209–12.
- 710 40. Venables WN, Ripley BD. *Modern applied statistics with S.* 4th ed. New York: Springer; 2002. (Statistics and computing).
41. Thiele M, Detlefsen S, Sevelsted Moller L, Madsen BS, Fuglsang Hansen J, Fialla AD et al. Transient and 2-Dimensional Shear-Wave Elastography Provide Comparable Assessment of Alcoholic Liver Fibrosis and Cirrhosis. *Gastroenterology* 2016; 150(1):123–33.
- 715 42. Karlas T, Dietrich A, Peter V, Wittekind C, Lichtinghagen R, Garnov N et al. Evaluation of Transient Elastography, Acoustic Radiation Force Impulse Imaging (ARFI), and Enhanced Liver Function (ELF) Score for Detection of Fibrosis in Morbidly Obese Patients. *PLoS ONE* 2015 [cited 2020 Sep 22]; 10(11):e0141649.
- 720 43. Pu K, Wang Y, Bai S, Wei H, Zhou Y, Fan J et al. Diagnostic accuracy of controlled attenuation parameter (CAP) as a non-invasive test for steatosis in suspected non-alcoholic fatty liver disease: a systematic review and meta-analysis. *BMC Gastroenterol* 2019 [cited 2020 Apr 21]; 19(1):51.
44. Caussy C, Brissot J, Singh S, Bassirian S, Hernandez C, Bettencourt R et al. Prospective, Same-Day, Direct Comparison of Controlled Attenuation Parameter With the M vs the XL Probe in Patients With Nonalcoholic Fatty Liver Disease, Using Magnetic Resonance Imaging-Proton Density Fat Fraction as the Standard. *Clinical Gastroenterology and Hepatology* 2020; 18(8):1842–50.
- 725 45. Shalimar, Kumar R, Rout G, Kumar R, Yadav R, Das P et al. Body mass index-based controlled attenuation parameter cut-offs for assessment of hepatic steatosis in non-alcoholic fatty liver disease. *Indian J Gastroenterol* 2020 [cited 2020 May 23]; 39(1):32–41.
- 730 46. Oeda S, Takahashi H, Imajo K, Seko Y, Ogawa Y, Moriguchi M et al. Accuracy of liver stiffness measurement and controlled attenuation parameter using FibroScan® M/XL probes to diagnose liver fibrosis and steatosis in patients with nonalcoholic fatty liver disease: a multicenter prospective study. *J Gastroenterol* 2020 [cited 2020 Apr 28]; 55(4):428–40.
47. Tandra S, Yeh MM, Brunt EM, Vuppalanchi R, Cummings OW, Ünalp-Arida A et al. Presence and significance of microvesicular steatosis in nonalcoholic fatty liver disease. *Journal of Hepatology* 2011 [cited 2020 Jul 15]; 55(3):654–9.
- 735 48. Younossi ZM, Golabi P, Avila L de, Paik JM, Srishord M, Fukui N et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: A systematic review and meta-analysis. *Journal of Hepatology* 2019; 71(4):793–801.
49. Ballestri S, Nascimbeni F, Lugari S, Lonardo A, Francica G. A critical appraisal of the use of ultrasound in hepatic steatosis. *Expert Rev Gastroenterol Hepatol* 2019; 13(7):667–81.
- 740 50. Angulo P, Bugianesi E, Bjornsson ES, Charatcharoenwitthaya P, Mills PR, Barrera F et al. Simple noninvasive systems predict long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology* 2013; 145(4):782.
- 745 51. Ajmera V, Park CC, Caussy C, Singh S, Hernandez C, Bettencourt R et al. Magnetic Resonance Imaging Proton Density Fat Fraction Associates With Progression of Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2018 [cited 2020 Sep 17]; 155(2):307-310.e2.
52. Arulanandan A, Ang B, Bettencourt R, Hooker J, Behling C, Lin GY et al. Association Between Quantity of Liver Fat and Cardiovascular Risk in Patients With Nonalcoholic Fatty Liver Disease Independent of Nonalcoholic Steatohepatitis. *Clin Gastroenterol Hepatol* 2015 [cited 2020 Sep 17]; 13(8):1513-20.e1.
- 750 53. Runge JH, Smits LP, Verheij J, Depla A, Kuiken SD, Baak BC et al. MR Spectroscopy-derived Proton Density Fat Fraction Is Superior to Controlled Attenuation Parameter for Detecting and Grading Hepatic Steatosis. *Radiology* 2018 [cited 2020 Apr 22]; 286(2):547–56.
54. Imajo K, Kessoku T, Honda Y, Tomeno W, Ogawa Y, Mawatari H et al. Magnetic Resonance Imaging More Accurately Classifies Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease Than Transient Elastography. *Gastroenterology* 2016; 150(3):626-637.e7.
- 755 55. **Harry S, Lai LL**, Nik Mustapha NR, Abdul Aziz YF, Vijayanathan A, Rahmat K et al. Volumetric Liver Fat Fraction Determines Grade of Steatosis More Accurately Than Controlled Attenuation Parameter in Patients With Nonalcoholic Fatty Liver Disease. *Clin Gastroenterol Hepatol* 2020; 18(4):945-953.e2.

- 760 56. Patterson B, Miller DL. Experimental Measurements of Ultrasound Attenuation in Human Chest Wall and Assessment of the Mechanical Index for Lung Ultrasound. *Ultrasound in Medicine & Biology* 2020; 46(6):1442–54.
57. Fujiwara Y, Kuroda H, Abe T, Ishida K, Oguri T, Noguchi S et al. The B-Mode Image-Guided Ultrasound Attenuation Parameter Accurately Detects Hepatic Steatosis in Chronic Liver Disease. *Ultrasound in Medicine & Biology* 2018 [cited 2020 Sep 17]; 44(11):2223–32.
- 765 58. Ferraioli G, Maiocchi L, Raciti MV, Tinelli C, Silvestri A de, Nichetti M et al. Detection of Liver Steatosis With a Novel Ultrasound-Based Technique: A Pilot Study Using MRI-Derived Proton Density Fat Fraction as the Gold Standard. *Clin Transl Gastroenterol* 2019 [cited 2020 May 4]; 10(10):e00081.
59. O'Sullivan JW, Banerjee A, Heneghan C, Pluddemann A. Verification bias. *BMJ Evid Based Med* 2018 [cited 2020 May 7]; 23(2):54–5.
- 770 60. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ* 2016 [cited 2020 May 15]; 353:i3139.
61. Pournik O, Alavian SM, Ghalichi L, Seifizarei B, Mehrnoush L, Aslani A et al. Inter-observer and Intra-observer Agreement in Pathological Evaluation of Non-alcoholic Fatty Liver Disease Suspected Liver Biopsies. *Hepat Mon* 2014 [cited 2020 Mar 12]; 14(1):e15167.
- 775 62. **Eslam M, Newsome PN**, Anstee QM, Targher G, Gomez MR, Zelber-Sagi S et al. A new definition for metabolic associated fatty liver disease: an international expert consensus statement. *Journal of Hepatology* 2020 [cited 2020 Apr 25]; 73(1):202–9.
- 780 63. Newsome PN, Sasso M, Deeks JJ, Paredes A, Boursier J, Chan W-K et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. *The Lancet Gastroenterology & Hepatology* 2020 [cited 2020 May 4]; 5(4):362–73.

Author names in bold designate shared co-first authorship.

FIGURE CAPTIONS

785

Figure 1: Flowchart of the papers and patients included in the final analysis including probe data availability. The letters 'i' and 'n' represent the number of studies and patients respectively. CAP = controlled attenuation parameter.

790

Figure 2: Probability of steatosis grade as a function of controlled attenuation parameter (CAP) for non-alcoholic fatty liver disease (NAFLD) patients. The prevalence of each grade (S0–S3) is shown in brackets. The choice for each panel is based on the observed values from: (a) Eddowes 2019,(14) (b) Naveau 2017,(23) (c) Chan 2018,(26) (d) Baumeler 2019,(15) respectively.

795

Figure 3: The areas under the curve from the current analysis and the papers that comprise its data are ordered by aetiology. Note that Lai only contributed 26 patients not in the Chan paper. Cardoso 2019 had no S0 patients. ALD = Alcoholic liver disease; IPDMA = Individual patient data meta-analysis; NAFLD = Non-alcoholic fatty liver disease; ROC = Receiver operating characteristics.

Footnote to Figure 3

800

*In Eddowes 2019, only patients with suspected NAFLD were included. In 30 of 47 S0 patients, another diagnosis was established.